Table of contents

Part I: Building and Visualizing the Networks

The analysis will be using 2 types of files: "filtered_twitch_edges.csv" as the *base_network* and "filtered_twitch_features.csv" will make up the *node_attributes*.

```
 [1] 0.2205 0.0620 0.1619 0.5665 0.7603 0.0352 0.0015 0.1247 0.7477 0.0646
[11] 0.1156 0.1988 0.1195 0.3688 0.5924 0.0492 1.2423 0.0620 1.0746 0.1123
[21] 0.1112 0.2965 0.5065 0.5817 0.1913 0.5386 0.1206 0.0735 0.0363 0.9647
[31] 0.0242 0.6411 0.2895 8.0194 0.1548 0.0581 0.1721 0.0334 0.2291 0.0650
[41] 1.7945 0.0160 0.2179 0.4435 3.8916 0.1559 0.0564 3.2941 0.2292 0.0006
[51] 0.0488 0.3311 0.0334 0.2310 0.4249 1.3279 0.5163 0.4391 1.7174 0.0076
[61] 0.0376 0.2412 0.5223 3.5732 0.1277 0.2965 0.5710 0.2887 0.5580 0.3039
[71] 0.2095 0.1543 0.3703 0.0831 0.0323 0.9448 0.2356 0.0417 0.3579 1.2389
[81] 0.1186 0.0809 0.2460 0.4283 0.0586 0.7446 0.6424 0.0003 0.6086 0.1979
[91] 0.1389
```

```
 [1] 1 1 0 1 1 0 0 0 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1 0 0 0 1 0 0 1 1 1 0 0 0 0 0 0
[39] 0 0 0 1 0 1 1 0 0 0 0 0 1 1 0 0 1 1 0 0 1 0 0 0 1 1 0 1 1 1 0 0 1 1 1 1 0 1
[77] 0 1 1 1 0 0 1 1 0 0 1 0 1 1 1
```

```
 [1] 193 151 177 157 181 176  92 136 174 167 179 180 184 168 176 180 173 114 122
[20] 106 173 182 183 188 189 193 186 148 151 180 175 174 180 188 168 172 184 183
[39] 175 153 167 185 175 164 194 185 121 150 165  98 188 181 153 186 187 179 182
[58] 194 175 148 188 169 181 158 177 184 162 194 172 158 175 186 181 149 175 182
[77] 177 120 178 172 189 190 180 189 176 187 190  74 191 186 153
```

```
 [1] 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[39] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[77] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

```
 [1] "EN" "EN" "EN" "EN" "EN" "EN" "EN" "PT" "EN" "EN" "EN" "EN" "EN" "EN" "EN"
[16] "EN" "FR" "EN" "EN" "EN" "EN" "EN" "EN" "EN" "EN" "PT" "EN" "EN" "EN" "EN"
[31] "EN" "EN" "EN" "ES" "FR" "EN" "EN" "EN" "EN" "EN" "EN" "EN" "EN" "EN" "EN"
[46] "EN" "EN" "EN" "EN" "EN" "EN" "EN" "EN" "EN" "EN" "EN" "EN" "PT" "EN" "EN"
[61] "EN" "EN" "EN" "EN" "EN" "EN" "EN" "EN" "EN" "EN" "EN" "EN" "EN" "EN" "EN"
[76] "EN" "EN" "EN" "EN" "EN" "EN" "ES" "EN" "EN" "ES" "EN" "EN" "EN" "EN" "EN"
[91] "EN"
```

```
 [1] 1 1 0 1 1 1 0 0 1 1 1 1 1 1 1 0 0 0 1 0 1 1 1 1 1 1 1 0 1 1 0 1 1 0 0 0 1 0
[39] 1 1 1 0 1 1 1 1 1 1 1 0 0 1 0 1 1 1 1 1 1 0 0 1 1 0 1 1 1 1 1 1 1 1 1 0 0 1
[77] 1 0 1 1 1 0 1 1 0 1 1 0 1 0 1
```

Hypotheses:

We are going to test the following hypotheses in this project.

Hypothesis 1: Users who share the same primary language are more likely to form mutual follower connections than users with different primary languages.

Hypothesis 2: Users who are both affiliates are more likely to form mutual follower connections than users with different affiliate statuses.

Hypothesis 3: Users who have marked their accounts as mature are more likely to form mutual follower relationships with other users who have also marked their accounts as mature.

Hypothesis 4: Users with a higher number of views are more likely to form mutual follower connections than users with fewer views.

Hypothesis 5: Users with older accounts (higher life_time values) are more likely to form mutual follower connections.

Hypothesis 6: If user A and user B have a mutual follower relationship, and user B and user C also have a mutual follower relationship, then user A and user C are more likely to form a mutual follower relationship (i.e., the network exhibits a high degree of transitivity).

Hypothesis 7: Users are likely to form connections that create "2-star" configurations, where one user is connected to two others who are not connected to each other.

Hypothesis 8: Users are likely to form connections that create "2-star" configurations, where one user is connected to two others who are not connected to each other.

1. Plotting the mutual followers network.

```
Attaching package: 'igraph'


The following objects are masked from 'package:sna':

    betweenness, bonpow, closeness, components, degree, dyad.census,
    evcent, hierarchy, is.connected, neighborhood, triad.census


The following objects are masked from 'package:network':

    %c%, %s%, add.edges, add.vertices, delete.edges, delete.vertices,
    get.edge.attribute, get.edges, get.vertex.attribute, is.bipartite,
    is.directed, list.edge.attributes, list.vertex.attributes,
    set.edge.attribute, set.vertex.attribute
```

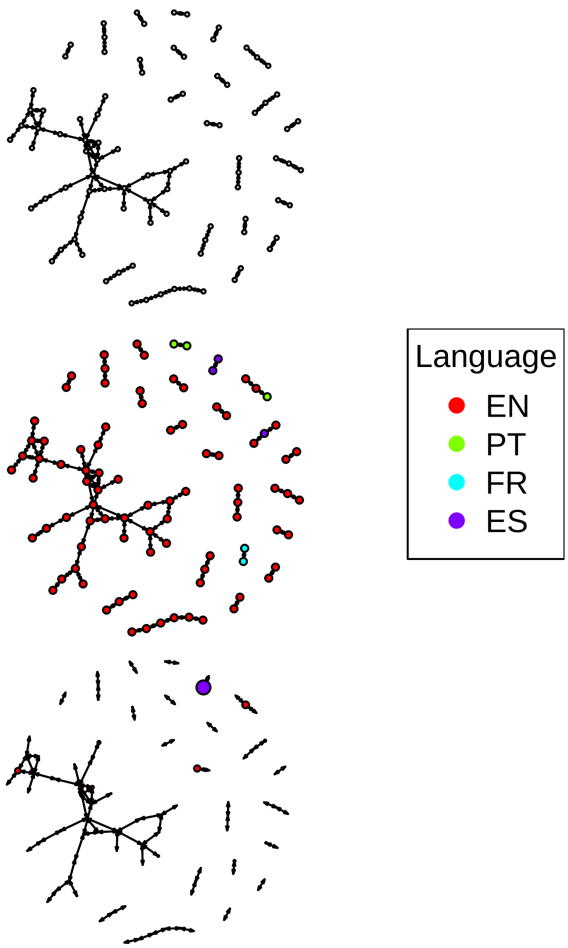
```
The following objects are masked from 'package:stats':

    decompose, spectrum

The following object is masked from 'package:base':

    union

Warning: `graph.adjacency()` was deprecated in igraph 2.0.0.
i Please use `graph_from_adjacency_matrix()` instead.
```



Part II: Model Estimation

```
Loading required package: slam

Using the GLPK callable library version 4.65

edges
   76
```

```
nodematch.mature
              38

nodematch.affiliate
                 56

nodematch.language
                73

kstar2
   110

nodecov.descaled_views
               84.9234

nodecov.life_time
            26113

gwesp.fixed.0.693147180559945
                         17.5

Call:
ergm(formula = mutual ~ edges + nodematch("mature") + nodematch("affiliate") +
    kstar(2) + nodecov("descaled_views") + nodecov("life_time") +
    gwesp(log(2), fixed = T), control = control.ergm(seed = 42),
    verbose = F)

Monte Carlo Maximum Likelihood Results:

                               Estimate Std. Error MCMC % z value Pr(>|z|)
edges                         -4.768539   1.455512      0  -3.276  0.00105 **
nodematch.mature              -0.027587   0.230011      0  -0.120  0.90453
nodematch.affiliate            0.682515   0.270202      0   2.526  0.01154 *
kstar2                        -0.353767   0.124504      0  -2.841  0.00449 **
nodecov.descaled_views         0.022410   0.091696      0   0.244  0.80692
nodecov.life_time              0.003344   0.004364      0   0.766  0.44350
gwesp.fixed.0.693147180559945  1.032190   0.180382      0   5.722  < 1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 5676.9  on 4095  degrees of freedom
 Residual Deviance:  723.9  on 4088  degrees of freedom

AIC: 737.9  BIC: 782.1  (Smaller is better. MC Std. Err. = 0.3712)
```

Interperating the Hypothesis

Hypothesis 1: Homophily on Language

The p-value for nodematch.language is:

0.002406

This is less than 0.05, meaning we reject the null hypothesis. There is likely a correlation between sharing the same language and forming a mutual follower relationship.

Now we look at the log-odds ratio:

1.957582

When exponentiated, we get:

7.085553504126308

This means that the formation of mutual follower relationships between users who share the same language is about 7 times more likely than between users who do not share the same primary language. Therefore, this hypothesis is accepted.

Hypothesis 2: Homophily on Affiliate Status

The p-value for nodematch.affiliate is:

0.033661

This is less than 0.05, so we reject the null hypothesis. There is likely a correlation between sharing the same affiliate status and forming a mutual follower relationship.

Now we look at the log-odds ratio:

0.566721

When exponentiated, we get:

1.7622939521970955

This indicates that the formation of mutual follower relationships between users who share the same affiliate status is about 1.76 times more likely than between users who do not share the same affiliate status. Therefore, this hypothesis is accepted.

Hypothesis 3: Homophily on Maturity

The p-value for nodematch.mature is:

0.753290

This is greater than 0.05, so we cannot reject the null hypothesis. The data does not provide sufficient evidence to support a correlation between sharing the same maturity status and forming a mutual follower relationship.

Hypothesis 4: Influence of View Count on Connectivity

The p-value for nodecov.descaled_views is:

0.271689

This is greater than 0.05, meaning we cannot reject the null hypothesis. The data does not provide sufficient evidence to support the hypothesis that having a higher number of views increases the likelihood of forming a mutual follower relationship.

Hypothesis 5: Effect of Account Age on Connectivity

The p-value for nodecov.life_time is:

0.300042

This is greater than 0.05, meaning we cannot reject the null hypothesis. The data does not provide sufficient evidence to support the hypothesis that users with older accounts are more likely to form mutual follower relationships.

Hypothesis 6: Transitivity in the Network

The p-value for gwesp.fixed.0.693147180559945 is:

$< 1e\text{-}04$

This is less than 0.05, indicating we reject the null hypothesis and conclude that transitive relationships are likely correlated with the formation of mutual follower relationships.

Now we look at the log-odds ratio:

1.010757

When exponentiated, we get:

2.7474177409933193

This suggests that if user A and user B have a mutual follower relationship, and user B and user C also have a mutual follower relationship, then user A and user C are about 2.75 times more likely to also have a mutual follower relationship compared to pairs of users who do not have such transitive relationships. Therefore, this hypothesis is accepted.

Hypothesis 7: Role of 2-Star Configurations

The p-value for kstar2 is:

0.000968

This is less than 0.05, so we reject the null hypothesis. There is a significant correlation between the presence of 2-star configurations and the formation of mutual follower relationships.

Now we look at the log-odds ratio:

-0.439450

When exponentiated, we get:

0.6441856253239544

This indicates that the presence of 2-star configurations decreases the likelihood of forming mutual follower relationships. Specifically, users connected through 2-star configurations are about 0.64 times as likely to form mutual follower relationships compared to users who are not part of such configurations.

Interpretation: The data does not support the hypothesis that 2-star configurations increase the likelihood of mutual follower relationships. Instead, it suggests the opposite: mutual follower relationships are less likely to occur in the context of 2-star configurations. This could indicate that users are more inclined to form fully interconnected relationships (triads) rather than being part of partially connected structures like 2-stars. Therefore, the original hypothesis is rejected.